Chapter 1

Introduction

Why is performance testing important and why do it in the first place? Badly performing applications are not a great benefit to a company or an organization. These applications mostly create a net cost of time and, of course, money. Moreover, if application does not deliver benefits, its future is not bright.

Performance testing is very significant, but underrated part of testing (unlike unit or functional testing). Sadly, it is not appreciated among executives and its importance is, as already mentioned, ignored. This fact has slightly changed in the last decade despite efforts of many known software consultants and highly publicized failures of key software applications.

Chapter 2

Python-nitrate

python-nitrate is a Python interface to the Nitrate test case management system. The package contains low-level driver (which allows direct access Nitrate's xmlrpc API), a high-level Python module (natural object interface) and a command line interpreter (useful for fast debugging and exprimenting).

This is an open source project with accessible source codes and RPMs that can be viewed and downloaded here:

```
http://psss.fedorapeople.org/python-nitrate/download/
```

2.1 Caching in python-nitrate

python-nitrate client supports currently 4 types of caching. First type is **CACHE_NONE**, where caching is disabled and every object change is immediately written to server. Every query is sent to server even if several same queries entered in a row. Second type is **CACHE_CHANGES**. When using this caching type, caching is enabled and changes are pushed to the server only if operation *update()* is entered or upon destruction. Third type is **CACHE_OBJECTS**, where any loaded object is saved for possible future use. Finally, caching type **CACHE_ALL** caches all available objects (those which are in class, for example if user name is required with ID 123, every user will be cached).

Caching type			
CACHE_NONE	**CACHE_CHANGES**	**CACHE_OBJECTS**	**CACHE_ALL**
download download	download and store use cached	download and store use cached	download all and store use cached

Table 2.1: Caching objects

2.2 Classes in python-nitrate

There are several classes in python-nitrate module.

	Caching type			
Class	**CACHE_NONE**	**CACHE_CHANGES**	**CACHE_OBJECTS**	**CACHE_ALL**
Bug	X	X	X	X
Build	OK	N/A	X	X
CaseStatus	N/A	N/A	N/A	N/A
Category	OK	N/A	OK	X
CaseRun	OK	OK	X	X
TestCase	OK	OK	X	X
TestPlan	OK	OK	X	X
TestRun	OK	OK	X	X
PlanStatus	N/A	N/A	N/A	N/A
PlanType	OK	N/A	OK	X
Priority	N/A	N/A	N/A	N/A
Product	OK	N/A	X	X
RunStatus	N/A	N/A	N/A	N/A
Status	N/A	N/A	N/A	N/A
User	OK	OK	OK	OK
Version	OK	N/A	X	X

Table 2.2: Classes and their type of caching

Chapter 3

Nitrate

Nitrate is a test case management system (TCMS), it is written in programming language Python and uses Django web framework. This management system provides several features:

- Python interface (concise and natural)
- custom level of caching and logging
- status coloring
- integrated test suite
- utility functions
- multiple authentication backends
- XMLRPC interface
- audit traceability
- increased productivity – identification of gaps in product coverage

Hiding unnecessary implementation details wherever possible was the main motivation resulting in concise use of the API.

Nitrate is an open source project and its source code can be viewed via browser at:

```
https://git.fedorahosted.org/git/nitrate.git
```

or using git:

```
git clone git://git.fedorahosted.org/nitrate.git
```

Chapter 4

What is performance?

Performance of application depends on perception. A well-performing application application lets the user carry out a given task without irritation or perceived delay. Performant application does not display blank screen during login and can complete user's task without letting their attention wander. However, delivering acceptable level of performance is a struggle for a lot of applications.

In this context, application is being referred to as a whole, since it is composed of many parts. The higher level consists of the application software and the application landscape. For example servers required to run and also the network infrastructure for communication are latter. Either way, if problems occur in any of these areas, application performance is sure to decline.

Some people may say the best approach to ensuring good application performance is to observe the behavior of each of these areas and (under load and stress) and solve any problem that occurs. This is a common mistake. since you end up dealing with the symptoms of performance rather than dealing with the cause.

4.1 Measuring the performance

There are several key indicators that must be taken into account. They could be divided into two groups: service-oriented and efficiency-oriented. **Service-oriented** indicators measure how well (or not) an application is providing a service to the users. These indicators are availability and response time. **Efficiency-oriented** indicators measure how well (or not) an application makes use of the application lanscape. These indicators are throughput and utilization.

These terms could be defined as following: Availability is the amount of time of application's availability to the user. This is very crucial, since lack of availability could have substantial business cost even for a small outage. In other words, this would mean complete inability for user to make effective use of the application.

Response time is amount of time that takes the application to respont to user request. Measuring system response time, the time between users request for response and from the application and a complete reply to the user, is sufficient for performance testing.

Throughput is the rate at which certain application-oriented events occur. For example the number of hits on a web page in a given period of time.

Utilization is the percentage of the teoretical capacity of a resource that is being used. For example the amount of memory used when certain number of visitors is present on a server or how much bandwidth is being consumed by traffic of application.

4.1.1 Performance standards

Although there have been attempts to define a standard (particularly for web-based applications), there is no generic standard to evaluate if performance of application is good or bad. Good example is the minimum page refresh time. It came from 20 seconds rapidly to 8 seconds, but the application user wants instant response, which is likely to remain elusive.

There was a research in 1988 that attempted to map user productivity to response time. Even though the research was based on green-screen text applications, its conclusions are probably still relevant.

Conversational interaction should not be greater than 15 seconds. Otherwise waiting for the answer becomes intolerable for a busy call-center operator of futures trader. The system should be designed to allow the user turn to other activities if such situation occurs.

Delays greater than 4 seconds are too long for the user to retain information in short-term memory. These could inhibit problem-solving activity and data entry. But after the transaction is completed, delays between 4 – 15 seconds are acceptable.

For operations demanding high level of concentration, delays should not exceed 2 seconds. Waiting between 2 to 4 seconds can seem surprisingly long when the user concentrates and is emotionally commited to completing the task at hand. Response time should be than 2 seconds when the application user has to remember information through several responses. With the level of detailed information rises the need for responses shorter than 2 seconds. When work is thought-intensive (for example writing a book), in order to maintain the users' interest very short response time is required (under 1 second). Response time to pressing a key (and displaying it) should not be greater than 0.1 seconds, as well as clicking on object with mouse. Moreover, applications like computer games require extremely fast interaction. To sum up, critical response time seems to be 2 seconds. If the delays are greater, user's productivity will almost certainly decline. Obviously, the nominal page refresh time is not ideal since 8 seconds is highly above the critical barrier.

!Why is bad performance so common

The IT Business Valve Curve

It is a common habit that performance problems surface late in the application life cycle. There is a rule: the later you discover them, the greater will be the cost to resolve (as you can see on valve curve).

In this figure expected outcome is represented by solid line and planned moment of deployment by black diamond. If the application is released on schedule and immediately starts to provide benefit to the business with nearly no problems after realase. The dotted line represents reality when development and deployment target slip, which is portrayed as striped diamond and it involves great effort to fix performance issues in production costing a lot of time and money.

4.1.2 What the analysts think

Forrester Research in 2006 provided interesting results by looking at the number of performance defects that had to be fixed in production for a typical application deployment. The results could be viewed in this table:

Three levels of performance were identified. The first level is firefighting, which is typical for applications where nearly no or just little performance testing was carried out in application deployment. This means, that all performance problems have to be resolved in live environment (after the application is released). Although this approach is the least desirable, it is quite common and puts companies into serious risk. The second level is performance validation and it covers companies that

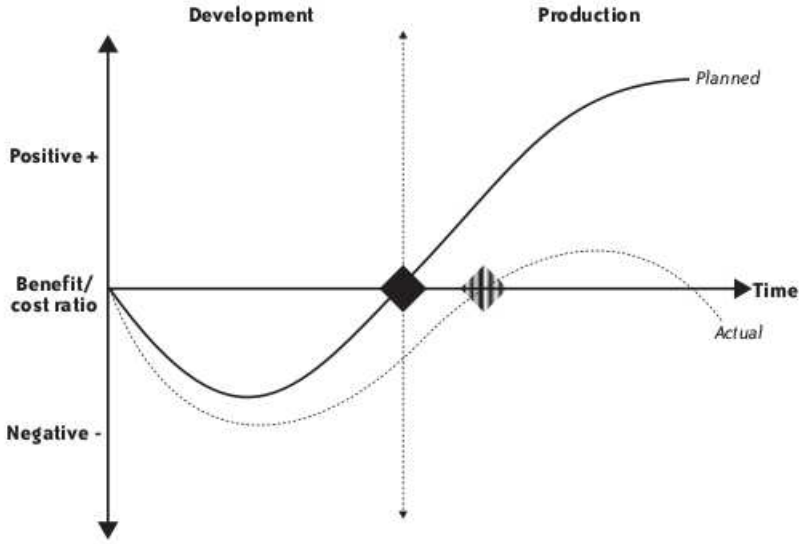

Figure 4.1: The IT Business valve curve

Approach	% resolved in production
Firefighting	100%
Performance Validation	30%
Performance Driven	5%

Table 4.1: Resolving performance defects (2006)

spent time by performance testing only in the late phase in the application life cycle. Despite this fact, still rather large number of performance defects (30%) are present in productiton. This level is the most common approach for organizations. Finally, performance driven approach is a stage where performance testing has been conducted at every stage of the application life cycle. The outcome of this is that only small percent of performance problems are discovered after deployment (only 5%). This should be aim for every company to adopt it as a good habit.

Last minute performance testing

Even though performance validation mode is much better than firefighting mode, last minute performance testing is still pretty dangerous. There is always a risk that serious performance defects that surface in production and the time to correct problems identified before release won't be sufficient. This could cause delays in application rollout, or application is deployed with severe performance defects and will cost a lot of time and money. Moreover, application might be withdrawn from circulation until problems are fixed. These outcomes affect negatively the business and the confidence of the users of the application. The best approach is to test for performance as soon as possible instead than postpone it to last minute.

Another unfortunate habit is that developers and testers overlook the amount of people in user community and their geography (developers are likely to ignore large number of users who have low-bandwidth, high-latency WAN links. Underestimating the popularity of (mostly) web application is no strange thing either. For example, if you estimate 10000 hits on your new web page and it suddenly becomes 1 million hits, your application infrastructure will most likely collapse.

4.2 Fundamentals of effective application performance testing

Performance awareness should be built into the application life cycle as early as possible

Performance requirements can be divided into functional an nonfunctional requirements. Regression and unit testing are considered as functional, performance related requirements can be considered as nonfunctional requirements. These are very important when trying to carry out effective performance testing.

Performance testing is much more than generating load and seeing what happens. Many other factors must be taken into account before appropriate performance testing strategy can be implemented. According to book TAOPT, these are the most important requirements for performance testing:

- choose appropriate performance testing tool
- design appropriate evironment for performance testing
- setting performance targets that are realistic and appropriate
- application has to be stable for performance testing
- obtaining a code freeze
- scripting and identifying critical transactions for business
- high quality test data
- accurate performance test design
- identification of server and network mointoring key Performance Indicators (KPIs)
- engough time to performance test effectively

Although a lot of these requirements are obvious, some of them are not. Generally, it's the requirements you underestimate have actually the gratest impact on the success or failure of performance testing. Every point will be examined in detail.

4.2.1 Choose appropriate performance testing tool

During the last 15 years, automated tools transformed from "fat-client" norm to web-enablement. This norm is better for performance tester because of more automated tools vendors to choose from, thus tester can choose from offering with even low budget (open source tools are available as well on http://www.opensource.org/). However, if the needs of performance testing move outside the Web, the number of available tools decreases and obsolete (and bad) technologies are still present. These technologies center on recording application activity and modifying resulting scripts for performance test rather than execution and analysis. Web-based technologies can cause some problems for performance testing tools too. For example, not all tools will be able to offer a solution when dealing with streaming media or client certificates. Despite these problems, automated tools are required in order to carry out serious performance testing. There is no practical way to perform reliable (and repeatable) performance testing without some form of automation.

The main goal of automated performance testing tools is to simplify the testing process. Normally, automated tools record user activity and render this data as transactions or scripts. These scripts are

later used to create scenarios that represent a mix of typical user activity or to create load testing sessions and can be considered as actual performance tests. Once created, they can easily be reused, which is a great advantage compared to manual testing. One of the greatest advantages over manual testing is the option to correlate performance data from various sources (for example the network, servers, applicaiton response time) and display them in a single view. This information is stored for each test run, thus making comparison of the multiple results easy.

4.2.2 Testing tool architecture

Typical components of an automated perfromance test tools are:

Scripting module enables recording of user activity and possibly support many middleware protocols. Modification of the scripts should be allowed to associate internal and external data and to configure granularity of response-time measurement.

Test management module is responsible for the creation and execution of load tests sessions or, for example, scenarios that represent different mixes of user activity. The sessions use scripts and one or more *load injectors* (which generate the load, it could generate from multiple workstations or servers, depending on the amount of load that is required).

Analysis module has the ability to analyze the data collected after test is executed. Obtained data is generally a mixture of autogenerated reports and the report is in graphical or tabular form. Also, an 'expert' capability could be present (automated analysis of results and point out areas of concern).

Monitoring network and server performance while load is test is running is possible with additional modules. In Figure 2 is a demonstration of a typical performance tool deployment. Group of servers or workstations will inject application load and will represent application users by creating 'virtual users'.

4.2.3 Designing an appropriate performance test environment

Ideally, the best test environment would be an *exact copy* of the deployment environment, but this case is very rare (for a number of reasons). Therefore the typical performance test environment could be a *subset of the deployment environment*. But you should certainly attempt to make the performance test environment as close to a replica of the live environment as possible within existing constraints. This is different from unit testing, where the goal is to ensure that the application works correctly.

It could take several weeks or even months to set up acceptable performance test environment, since the process is rather difficult. Therefore, to complete this activity, you need to plan for a realistic amount of time. Understanding the entire test environment enables more efficient test design and planning. Moreover, it can help to identify testing challenges early in the project. Occasionaly, this process must be revisited periodically throughout the life cycle of the project.

Virtualization

Virtualization is a relatively new factor influencing design of test environment, which allows multiple 'virtual' servers to exist on a single physical machine. Ideal conditions are when live environment also makes use of virtualization. If so, very close approximation will be possible between live and test enironment.

Injection Capacity

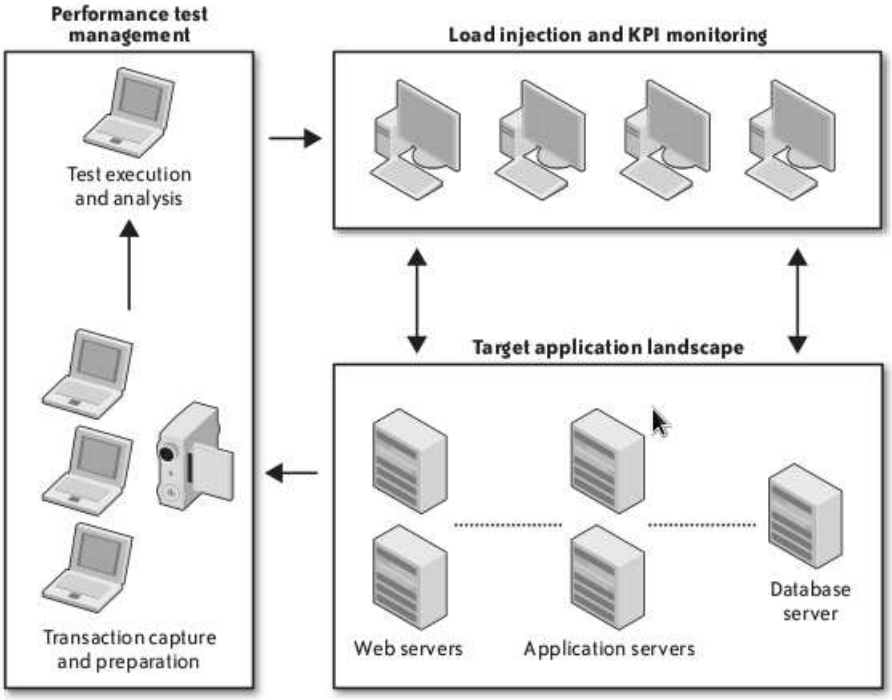

Figure 4.2: Performance tool deployment

In order to generate the required load for performance testing, you need to ensure that hardware resources are sufficient. One or more machines are used to simulate real user and to generate load using automated performance test tools. Of course, there is a limited number of users you can generate from machine (it depends on technology). Another important task is to make sure that none of the injectors are overloaded (CPU or memory utilization), because it may have severe impact on results of the performance test. Representing many users with only one machine is a compromise in automated performance testing, so the goal should be to use as many machines to represent load injection as possible (and thus spreading the load).

4.2.4 Setting Realistic and Appropriate Performance Targets

Performance targets are often referred to as performance goals or a Service Level Agreement (SLA). It is crucial to have clearly defined performance testing targets, otherwise it may be a waste of time.

Consesus

Consensus on the performance targets is critical, so everyone involved and on whom the application will have an impact, from application users to senior management must agree on the same performance targets.

Unfortunatelly, promoting consensus have not been important, since performance testing has been always last-minute activity or even completely omitted. Gaining consensus on performance testing

projects within an organization should contain promoting a culture of consultation and involvement. Therefore, interested parties should be involved in the project at its early stage.

The following groups or individuals should be ideally involved:

- Chief Information Officer (CIO)
- Chief Technology Officer (CTO)
- Chief Financial Officer (CFO)
- Departmental Heads
- The developers
- The testers (could be internal or outsourced)
- The infrastructure team(s)
- The end users

Key Performance Targets

Generally, there are three performance targets that apply for any performance testing. These are based on service-oriented performance indicators:

- Availability or uptime
- Concurrency, scalability and throughput
- Response time

The following, which are as much a measure of capacity as of performance, can be added:

- Network utilization
- Server utilization

Availability or uptime

This is very simple requirement: the application has to be available to the user at all times, the only exception is planned maintenance. It certainly must not fail within the target level of throughput or concurrency. Testing the availability is not as simple as it seems. For example, a successful ping of the server's physical machine does not necessarily mean the application is available. Another important parameter here is load. The application might run very well at modest loads, but when load increases, it may start to time out and return errors, thus suggesting lack of capacity for the generated load.

Concurrency, scalability and throughput

First, the defifnition of the word *concurrency* is required. According to Scott Barber's white paper [1], concurrency is (from the perspective of a performance testing tool) the number of active users generated by the software, which is not necessarily the same as the number of users accessing the application concurrently. The point is, that capacity goals can be derived from concurrency and scalability. Achieving the scalability targets demonstrates sufficient capacity in the application landscape

for the application to deliver to the business.

In terms of performance testing, two distinct areas are referred as *concurrency*:

Concurrent virtual users are understood as the number of active virtual users. This number is often different from the number of users actually accessing the testing application.

On the other hand, *concurrent application users* is the number of users that are currently accessing (are logged in) the tested application. This is key measure of how many virtual users are active in a certain moment. Another thing that needs to be decided is if the process of login and logout (and everything it involves) will also be a part of application activity testing. If we decide to include it to include these processes, users which are logged out will be not truly concurrent with other users. There are a lot of solutions to this problem (increase the execution time or persistence of each transaction, ...), but if these processes are not included as a part of testing, it becomes a whole lot easier.

The 80/20 rule applies (among many other things) also in performance testing: out of 100 users of application, around 20 users will be using the application anytime during the working day. Of course, allowances for usage peaks outside of normal limits should be included in testing. For example, in a large university, day-to-day usage could be relatively flat, but only at certain times of the year (student enrollment, published results, registration of projects, ...) concurrent usage increases significantly.

When the application is considered as stateless, the performance testing target is *throughput* rather than concurrency. This situation happens, when the application has no concept of the traditional logged-in user. This kind of performance is measured as 'hits' per minute or per second.

Response Time

As mentioned before, good response time is a matter of perception. (not complete)

Network Utilization

The impact of network utilization on performance testing depends on the available bandwidth between servers and the user. Exhausting available bandwidth is much less probable in a modern data center compared to in-house testing. However, when moving closer to the user side, change in performance can be significant - especially when communication involves the Internet. Large numbers of network conversations with high data presentations rates will have strongest impact on the transmission path with the lowest bandwidth.

Typical network metrics that should be measured while performance testing include the following: *Data volume* is the amount of data presented to the network. This is fairly important when users will be connecting to the application over low-bandwidth WAN links. High data volume combined with network latency effects and bandwidth restricitions does not usually yield good performance.

Data throughput is the rate that data is presented to the network. Performance target could be just several bytes per second. This target can be achieved by monitoring data throughput and it can discover if any problems are occurring. Unexpected reduction in data throughput is often the first indication of capacity problems, where the servers cannot keep up with the requests and virtual users start to suffer from time-outs.

Data error rate occurs when large number of network errors that require retransmission of data slow down throughput, thus degrading performance of the application.

Server utilization

Server resources that an application is allowed to use might be limited. This can be determined by monitoring KPIs while the server is under load. Many server performance metrics can be monitored, but the most important are:

- CPU utilization
- Memory utilization
- Disk input and output
- Disk space

(complete?)

4.2.5 Stable application for performance testing

When the test environment is provided and performance targets are set, application stability for performance testing has to be confirmed. This may seem obvious, but performance testing turns quite often to bug-fixing exercise and time for testing declines rapidly.

Stability of an application could be defined as the confidence that an application does exactly what it says on the box. If there are serious problems with the functionality of the application, there is no point in performance testing, since these problems will mask important problems that are results of stress and load. Naturally, code quality is significant factor in performance testing and is paramount to good performance.

There are several areas that can hide problems:

High data presentation might be a serious problem even though the application is functionally stable and it can be due to coding or design inefficiencies. User restricitions in bandwidth will certainly have negative impact on performance. Therefore application should not have redundant conversations between client and server, also excessive data (large images, . . .) within a web page are not preferred.

Poorly performing SQL is another problem, where if using SQL database, bad coding or configuration of stored procedures usually causes decline in application's preformance. These flaws have to be identified and corrected, since this effect on the performance testing will only be magnified under increasing load.

Large number of application network round trips would make application vulnerable to the effects of latency, bandwidth restriction and network congestion.

Undetected application errors may be also a problem. Even though the application works properly from a functional perspective, there might be errors that are not apparent to the users or the developers. They may be creating inefficiencies that can impact performance and scalability. For example, several HTTP 404 Not Found errors in a single transaction might not be a big problem, but multiply it by several thousands transactions per minute could have serious impact on performance.

4.2.6 Ensuring Accurate Performance Test Design

When key transactions and their data requirements are identified, the next step is creating number of different types of performance tests. There are several types of tests:

Baseline test is used to establish a point of comparison for further test runs (mainly for measuring response time of transaction). This test is executed as a single virtual user for a single transaction for a set period of time (or for set number of transaction iterations). There should be no other activity on the system involved, since the goal is to provide 'best case' measurement. Obtained value is used to determine the amount of performance degradation, in response to increasing number of users or throughput.

Load test is a standard performance test, in which the application is loaded up to the target concurrency. It is the closest approximation of real application use, including simulation of user interaction with the application client. Delays and pauses experienced during data entry are taken into account as well as responses to information returned from the application servers.

Stress test causes the application or some part of infrastructure to fail in order to determine the upper limits or sizing of the infrastructure. Stress test contiues until something breaks: response time exceeds the value specified as acceptable, no more user can log in or the application becomes unavailable. The point of this testing is when for example the target concurrency is 2000 users and the infrastructure fails at only 2005 users, it is good to know this because it shows that there is very little extra capacity available.

Soak or stability testing identifies problems that may appear only after extended period of time. The most common example is a slowly developing memory leak or an unpredicted limitation in the number of times that a transaction can be executed. Server monitoring is required when carrying out these type of tests. The problems will manifest typically as a gradual shutdown in response time or as an application's sudden loss of avaliability. In order to ensure accurate diagnosis, the correlation of data from users and servers at the point of failure is vital.

Smoke test focuses on what has changed. Therefore, smoke test should involve only those transactions that have been affected by a code change.

Isolation test usually consists of repeated executions of specific transactions that have been identified as a result of a performance issue.

Baseline, load and stress test should be always executed. Smoke test and soak test are more dependent on application and the time

4.3 Server and Network KPIs (Key Performance Indicators)

Identifying and onitoring server and network performance metrics should be done for the application. This is vital to achieve root-cause analysis of any problems that can appear while performance testing the application. In ideal situation, the automated peformance test solution should contain this monitoring Lack of integration is not an excuse for omitting this phase.

Server KPIs

Measuring the performance of server is done by monitoring software which observes the behavior of specific performance counters and metrics. In Unix/Linux world, these utilities are monitor, top, vmstat, iostat, SAR that monitors server KPIs. Monitorins using several layers is recommended, where the top layer is called generic monitoring, which focuses on a small amount of counters that will clearly tell if server is under stress. Next layers of monitoring should focus on specific technologies that are part of the application: such as the web, application and database servers.

There are several models or templates that can be used:

Generic templates is a common set of metrics that apply to all servers in the same tier having the same operation system. The purpose is to provide first-level monitoring of the effects of load and stress. These metrics are typically how busy the CPUs are and how much of memory is present.

These counters should be present in the template:

- Processor utilization percentage
- Top 10 processes
- Available memory in bytes
- Memory pages/second
- Processor queue length
- Context switches per second
- Physical disk: average disk queue length
- Physical disk: (Disk Time)
- Network interface: Packets Received Errors
- Network interface: Packets Outbound Errors

Database server tier

Most databases are similar in architecture, but differ from monitoring perspective. As a result, every database type will require its own template. Examples of database types:

- MySQL
- Oracle
- Microsoft SQL Server
- IBM DB2
- Sybase
- Informix

Network KPIs

While performance testing, packet round-trip time, data presentation and the detection of any errors that may occur as a result of high data volumes are the main focus of network monitoring. This capability might be built into automated performance test tool or it may be provided separately. However, if the guidelines on where to inject load were followed and the data presentation have been optimized, then network issues should be the least likely cause of problems during performance testing.

Available performance counters for Unix/Linux and Windows operating systems monitor the number of errors detected during a performance test execution as well as the amount of data being handled by each NIC card. Some automated performance test tools even separate server and network time for each element within a page to help differentiate between server and network problems.

4.4 Interpreting Results

It is vitally important to interpret the results of performance test correctly. Since proper performance target has been set as part of testing requirements, problems should be spotted quickly during the test or as part of the analysis at test completion. Another important thing to do is to have all the necessary information at hand for further diagnosis.

4.4.1 The Analysis Process

There are two approaches how to perform analysis: in real time (as the test executes) or at its conclusion.

Real-Time Analysis is basically waiting for something to happen or for the test to complete without apparent incident. When a problem occurs, monitoring tools are responsible for reporting the location of the problem in the application landscape.

4.4.2 Performance test output types

The root of performance test results analysis is statistical analysis. There are several types of statistical data that can be extracted from these results:

Mean is basically the average of a set of values. Response times are aggregated with mean to derive their average. Even though there exist many types of means, the most important for performance testing is 'arithmetic mean'. For example, the arithmetic mean of 3.5, 4.6, 4.0, 4.2 and 3.9 is 4.04. **Median** is the middle value in a set of numbers. It is not essential for performance testing results interpretation.

Standard deviation can be defined as variation or dispersion from the calculated average (mean) value. It is used to measure confidence in statistical conclusions. Data in random and real-life events tend to exhibit a **normal distribution** (also knows as the Bell Curve). Since high standard deviation most likely indicates erratic end user experience, an effort to achieve small standard deviation should be made. For example, when a mean response time is 60 seconds and standard deviation is 30 seconds, user has a high chance of experiencing response time from 45 seconds to 75 seconds.

[1]

Chapter 5

Practical part

5.1 Real-life use cases and test cases

5.1.1 Use Cases

Use Case 1: Updating Case Runs

Mass CaseRun update	**UC01**
Description:	Use case of updating states of Case Runs in a single Test Run
Culprit:	Updating status of multiple Case Runs is a very time consuming task when executed one-by-one (single update in one request)

Use Case 2: Print TestCase tags

Display Test Case Tags	**UC02**
Description:	Use case for displaying tags of a specified set of Test Cases
Culprit:	The main problem is that tags class are not implemented. When all test cases are downloaded (in one call), they are converted one-by-one (thus causing significant slowdown) and communication with server is redundant.

Use Case 3: TestCase not cached in CaseRun

TestCase not cached in CaseRun	**U0C3**
Description:	This use case shows test cases which are present in a specified test run
Culprit:	Test Cases linked to case runs are fetched every time from server (no use of cache)

Use Case 4: Search for test case and its owners

Search for test case and its owners	**UC04**
Description:	When searching for test cases, test case owners are requested
Culprit:	Main problem is fetching users from database. Search is only one query.

Use Case 5: Search TestCase plans

Search TestCase plans	**UC05**
Description:	Use case for displaying test cases from specified author and also test plans which contain these test cases
Culprit:	Test plans are downloaded twice (once in test case and then in test plan)

5.1.2 Test Cases

Test Case 1: Updating Case Runs

Covering mass CaseRun update use case is very important, because this is quite common task. The main problem is in the communication - every update is executed in a single call, instead of executing multiple updates in a single call. This feature is going to be very helpful especially for TestRuns with large number of TestCases.

Code:

```
for caserun in TestRun(self.performance.testrun):
    log.info("{0} {1}".format(caserun.id, caserun.status))
    caserun.status = Status(random.randint(1,8))
    caserun.update()
```

Basic description:

First of all, TestRun is fetched from server along with all its case runs. For every case run, its state is changed to random state (because update happens only when state is changed) and perform update of a single case run.

Test Case 2: Print TestCase tags

Use case 2 is base for this test case. Showing tags of multiple test cases in a test plan is really complicated. The main problem is that tags class are not implemented. When all test cases are downloaded (in one call), they are converted one-by-one (thus causing significant slowdown).

Code:

```
for case in TestPlan(self.performance.testplan):
    log.info("{0}: {1}".format(case, case.tags))
```

Basic description:

After initial fetch of test plan, it iterates through all of the test cases and gets id of test case's tag. This tag id is then converted into name using server call. This is very redundant, because one tag is fetched several times (more significant when a lot of test cases with the same tag).

Test Case 3: TestCase not cached in CaseRun

This test case shows the problem of handling test cases that occur frequently – for example linked to case runs. These case runs are linked to test runs and they are linked to test plans. Test runs within the same test plan are more or less similar, so caching test cases is really a good idea.
Code:

```
for testplan in TestPlan(self.performance.testplan).children:
    log.info("{0}".format(testplan.name))
    for testrun in testplan.testruns:
        log.info("  {0} {1} {2}".format(
                testrun, testrun.manager, testrun.status))
        for caserun in testrun.caseruns:
            log.info("    {0} {1} {2}".format(
                    caserun, caserun.testcase, caserun.status))
```

Basic description:

The first step is fetching children of specified test plan. In every child plan, test runs are iterated and in every test run, case runs are displayed. For every case run, their test case is called and it is displayed, along with caserun status.

Test Case 4: Search for test case and its owners

Test Case 5: Search TestCase plans

This covers use case when displaying test plans with their test cases and all test plans of test cases.

Problem is with fetching the same test plan more than once when multiple test cases contain the same test plan.

Code:

```
for testcase in TestPlan(self.performance.testplan):
    log.info("{0} is in test plans:".format(testcase))
    for testplan in testcase.testplans:
        log.info("  {0}".format(testplan.name))
```

5.2 Feature Enhancements

5.2.1 FE00: Test Suite

Performance test suite is the first step to application performance testing. During initial state of thesis, feedback from python-nitrate users was required and a lot of comments were received regarding performance issues. After processing these issues, five use cases that have to be improved in the terms of performance were created. These use cases are later converted into test cases, which performance test suite contains. Test cases can be found in the self test section of every affected class. The performance test suite can be run as:

```
# python api.py --performance [class]
```

Every performance test in test suite displays elapsed time in human readable format and the result of the test.
For running the performance test suite an additional section containing information about the test bed is required:

```
[performance]
testplan = 1234
testrun = 12345
```

Use the test-bed-prepare.py script attached in the test directory to prepare the structure of test plans, test runs and test cases. Nitrate test case management system instance is required to run this performance test suite.

5.2.2 FE01: MultiCall

MultiCall feature is used to encapsulate multiple calls to a remote server into a single request. If enabled, TestPlan, TestRun, TestCase and CaseRun objects will use MultiCall for updating their states (thus speeding up the process). Example usage:

```
multicall_start()
    for caserun in TestRun(12345):
        caserun.status = Status("IDLE")
        caserun.update()
multicall_end()
```

When multicall_start() is called, update queries are not sent immediately to server. Instead, they are queued and after multicall_end() is called, all queries are sent to server in a batch.
This feature enhancement resolves use case UC01, where CaseRun statuses are updated in one call instead of one call per one CaseRun update.

5.2.3 FE02: Tag Class

For handling tags of Test Cases, Test Runs and Test Plans, new class Tag is implemented. Tags are handled as objects instead being handled as strings, which provides much easier access and modification.

This is very important after implementing tags caching, because tags are stored in a single cache.

5.2.4 FE03: Common Caching

In order to save calls to server and time, caching support has been implemented. Every class that handles objects has its own cache and it is used to save already initialized (fetched) objects from server. Several classes are automatically fetched from server after initialization (immutable objects), the rest will be fetched from server upon request.

Currently, there are 3 types (levels) of caching:

```
CACHE_NONE - no caching at all
CACHE_CHANGES - caching only local updates of instance attributes
CACHE_OBJECTS - caching objects for further use (default setting)
CACHE_PERSISTENT - persistent caching (caching in a file) option enable
```

Persistent cache (local proxy) was another idea how to speed up performance of the module. It allows class caches to be stored in a file, load caches from a file, and clear caches. This performance improvement is very helpful mainly for immutable classes (for example User), where all user can be imported in the beggining of a script and a lot of connections can be saved.

This performance improvement can be activated only by specifying file name in config section ([cache]).

Cache expiration is a way how to prevent using probably obsoleted object (for example caserun). Every class has its own default expiration time, but, of course, it can be modified from config file (see example in [expiration]) and user input. Time unit in cache expiration is 1 hour.

There are two special values:

```
 0: EXPIRE_ALWAYS -> no caching of certain class
-1: EXPIRE_NEVER -> object never expires
```

Default values for specific classes:

The solution might be MultiCall feature. Solution can be implementing common caching class, so cached testcases can be used to improve performance. It can be improved by implementing Tag class with caching. One solution is persistent local cache with stored objects (users, ...). Solution might be caching test plans (so only one fetch is required) or persistent cache.

5.2.5 FE04: Persistent cache

Cached objects expiration

Since there are many types of objects / classes in python-nitrate module, every class has to have expiration. This means that after certain period of time, instance of class has to be fetched again from the server in order to stay up-to-date. Expiration times of classes differ, for example, immutable classes have much longer expirations than mutable objects. Here is table of expiration times in all classes:

5.2.6 FE05: Container initialization

This improvement is based on

5.3 Implementation and testing

All features listed in future features have been implemented into python-nitrate module in order to increase performance of this module. Development was incremental, that means after implementation of performance test suite and its integration, along with test-bed-prepare.py script, which creates needed structure in Nitrate instance, the focus was transferred to feature enhancements.

First enhancement, MultiCall, was implemented in the beggining and the initial results were more than satisfactory.

Performance testing of improvements is performed on server with Nitrate instance and three test plans are created. These test plans contain different number of test cases: 10, 100 and 1000 test cases. Every test plan has its child plan, which is basically the same test plan. One test run is created for every test plan.

The results of performance testing improvements are visible in Appendix A, where tables with times of all test cases are present.

10TCs	Test Case				
#	**TC01**	**TC02**	**TC03**	**TC04**	**TC05**
1.	31.38s	23.48s	14.94s	13.44s	45.24s
2.	32.66s	23.54s	16.77s	12.26s	44.91s
3.	29.84s	23.59s	17.15s	12.18s	44.77s
4.	35.41s	23.40s	16.73s	12.26s	44.72s
5.	38.27s	23.51s	16.80s	12.26s	43.84s
6.	35.53s	25.14s	16.67s	12.73s	44.39s
7.	32.82s	23.91s	16.96s	12.90s	44.66s
8.	35.48s	23.53s	17.05s	12.64s	44.59s
9.	35.58s	23.47s	16.94s	12.30s	45.43s
10.	29.79s	23.49s	16.73s	12.27s	44.39s
11.	32.59s	23.58s	16.78s	12.26s	44.65s
12.	30.40s	23.52s	16.76s	12.36s	45.19s
13.	30.13s	23.98s	16.77s	12.43s	44.85s
14.	24.37s	23.72s	16.76s	12.73s	44.57s
15.	27.20s	24.08s	17.79s	12.35s	45.17s
16.	29.79s	23.96s	16.81s	12.48s	44.90s
17.	29.91s	23.44s	16.80s	12.32s	44.72s
18.	35.50s	23.51s	16.76s	12.28s	44.97s
19.	36.10s	23.46s	16.95s	12.26s	44.44s
20.	27.48s	23.41s	16.79s	12.27s	43.88s
21.	35.49s	23.69s	16.80s	12.45s	44.59s
22.	32.60s	23.49s	17.00s	12.17s	44.25s
23.	35.55s	23.54s	16.69s	12.23s	45.46s
24.	29.71s	23.48s	16.80s	12.30s	44.36s
25.	29.77s	23.84s	16.84s	12.48s	45.10s
26.	35.88s	23.52s	16.80s	12.19s	44.61s
27.	35.47s	23.63s	16.88s	12.27s	43.96s
28.	32.81s	24.10s	16.75s	12.23s	44.19s
29.	35.43s	23.44s	16.79s	12.33s	44.71s
30.	35.55s	23.51s	16.78s	12.26s	44.60s
31.	35.55s	23.44s	17.12s	12.25s	44.71s
32.	26.80s	23.44s	16.76s	12.32s	43.85s
33.	26.99s	23.51s	17.19s	12.26s	50.15s
34.	33.05s	24.63s	16.92s	13.01s	44.69s
35.	30.03s	23.66s	16.71s	12.24s	44.34s
36.	32.53s	23.57s	16.81s	12.28s	44.27s
37.	27.01s	23.71s	16.85s	12.22s	45.14s
38.	35.44s	23.58s	16.74s	12.28s	44.42s
39.	35.61s	23.95s	16.91s	12.78s	44.98s
40.	33.17s	24.14s	17.03s	12.29s	45.58s
41.	33.41s	23.74s	17.03s	12.53s	45.27s
42.	27.04s	23.70s	16.76s	12.22s	44.38s
43.	32.52s	23.68s	16.74s	12.38s	45.45s
44.	32.58s	24.01s	16.93s	12.53s	45.32s
45.	35.32s	23.69s	16.83s	12.24s	44.15s
46.	26.87s	23.96s	17.12s	12.24s	44.49s
47.	33.13s	23.56s	17.19s	12.27s	45.02s
48.	32.72s	23.70s	16.78s	12.56s	44.99s
49.	35.87s	23.55s	16.76s	12.38s	45.16s
50.	29.83s	23.77s	16.79s	12.29s	44.26s

Table 5.6: Results of run with original implementation

100TCs	Test Case				
#	**TC01**	**TC02**	**TC03**	**TC04**	**TC05**
1.	387.32s	203.47s	26.92s	26.36s	537.17s
2.	410.50s	245.04s	32.15s	27.24s	488.32s
3.	381.52s	197.35s	26.20s	26.27s	476.62s
4.	375.46s	200.28s	26.56s	26.36s	473.17s
5.	377.97s	199.44s	26.11s	25.88s	469.89s
6.	385.58s	199.90s	26.43s	26.71s	473.94s
7.	366.94s	200.92s	26.29s	26.21s	473.82s
8.	370.27s	220.55s	26.27s	30.19s	510.02s
9.	384.30s	200.16s	27.44s	26.57s	477.51s
10.	379.99s	204.30s	26.08s	26.83s	483.67s
11.	397.09s	201.62s	26.37s	26.35s	479.87s
12.	389.61s	200.55s	27.32s	26.03s	472.05s
13.	393.08s	199.58s	26.43s	26.30s	475.68s
14.	367.32s	199.48s	26.11s	28.58s	478.47s
15.	373.31s	198.76s	26.72s	26.03s	471.10s
16.	391.41s	200.39s	25.96s	26.12s	472.16s
17.	399.34s	199.52s	26.42s	26.11s	479.01s
18.	384.74s	210.97s	26.81s	26.17s	472.88s
19.	365.13s	199.58s	26.04s	26.35s	474.90s
20.	365.73s	199.32s	25.97s	26.16s	472.93s

Table 5.7: Results of run with original implementation

10TCs	Test Case				
#	**TC01**	**TC02**	**TC03**	**TC04**	**TC05**
1.	15.17s	23.90s	42.23s	0.00s	112.40s
2.	13.02s	25.19s	42.26s	0.00s	121.19s
3.	14.13s	24.31s	42.06s	0.00s	117.27s
4.	14.64s	29.19s	42.90s	0.00s	98.272s
5.	13.98s	23.68s	41.52s	0.00s	101.67s
6.	14.95s	23.43s	41.51s	0.00s	98.521s
7.	18.64s	24.16s	41.52s	0.00s	97.526s
8.	11.44s	23.68s	41.42s	0.00s	97.701s
9.	12.97s	23.61s	41.57s	0.00s	97.988s
10.	14.12s	23.61s	41.64s	0.00s	97.747s
11.	14.96s	23.48s	41.62s	0.00s	99.564s
12.	13.03s	23.91s	41.45s	0.00s	100.25s
13.	14.11s	23.60s	42.60s	0.00s	97.842s
14.	13.22s	24.96s	42.49s	0.00s	99.893s
15.	11.48s	24.37s	42.16s	0.00s	99.963s
16.	14.16s	23.72s	41.89s	0.00s	98.834s
17.	15.10s	25.35s	43.10s	0.00s	105.80s
18.	10.53s	24.38s	44.62s	0.00s	116.31s
19.	14.94s	23.67s	42.41s	0.00s	114.44s
20.	14.01s	23.64s	41.48s	0.00s	114.85s
21.	14.08s	23.82s	42.85s	0.00s	111.51s
22.	13.03s	23.41s	42.77s	0.00s	111.25s
23.	11.34s	23.54s	41.36s	0.00s	100.05s
24.	13.03s	23.48s	41.17s	0.00s	98.225s
25.	13.04s	26.47s	42.55s	0.00s	97.650s
26.	14.04s	23.85s	42.15s	0.00s	96.986s
27.	13.42s	29.17s	53.79s	0.00s	97.544s
28.	14.02s	23.66s	41.43s	0.00s	98.086s
29.	14.98s	23.46s	41.51s	0.00s	97.438s
30.	11.38s	23.89s	41.48s	0.00s	97.184s
31.	11.38s	23.79s	41.49s	0.00s	97.539s
32.	14.05s	23.53s	42.05s	0.00s	102.49s
33.	13.07s	24.35s	41.34s	0.00s	98.271s
34.	11.45s	23.56s	41.44s	0.00s	97.616s
35.	13.93s	23.51s	41.41s	0.00s	97.681s
36.	15.01s	23.40s	41.58s	0.00s	97.101s
37.	14.07s	23.63s	42.76s	0.00s	99.172s
38.	13.08s	23.58s	41.34s	0.00s	97.714s
39.	13.08s	23.86s	41.41s	0.00s	97.779s
40.	12.94s	23.60s	41.61s	0.00s	100.17s
41.	14.03s	24.11s	42.24s	0.00s	98.390s
42.	13.22s	23.66s	41.74s	0.00s	98.584s
43.	13.01s	23.53s	41.48s	0.00s	98.254s
44.	13.98s	23.45s	41.56s	0.00s	97.987s
45.	14.01s	23.52s	41.53s	0.00s	98.570s
46.	11.47s	23.55s	47.32s	0.00s	97.669s
47.	14.05s	24.23s	41.47s	0.00s	98.068s
48.	13.22s	23.33s	42.10s	0.00s	98.076s
49.	13.99s	23.42s	41.44s	0.00s	111.50s
50.	14.72s	25.95s	47.01s	0.00s	118.63s

Table 5.8: Results of run with improved implementation (no persistent cache)

5.4 Results

The results are quite impressive. Here is a table of mean times of experiments:

	Test Case				
Type	**TC01**	**TC02**	**TC03**	**TC04**	**TC05**
Original implementation	32.28s	23.70s	16.83s	12.39s	44.81s
Improved implementation	13.57s	24.12s	42.35s	0.00s	101.90s
Improved + persistent cache	s	s	s	s	s

Table 5.9: Comparison of results with 10 Test cases

The obvious thing here is that even though there is great progress in first and fourth test case, regress occured in test case three and five in Improved implementation. This is because all immutable objects are immediately fetched and cached, whereas in the old implementation it was only initialization.

	Test Case				
Type	**TC01**	**TC02**	**TC03**	**TC04**	**TC05**
Original implementation	s	s	s	s	s
Improved implementation	s	s	s	s	s
Improved + persistent cache	s	s	s	s	s

Table 5.10: Comparison of results with 100 Test cases

	Test Case				
Type	**TC01**	**TC02**	**TC03**	**TC04**	**TC05**
Original implementation	s	s	s	s	s
Improved implementation	s	s	s	s	s
Improved + persistent cache	s	s	s	s	s

Table 5.11: Comparison of results with 1000 Test cases

5.5 Conclusion

The goal of this thesis has been improvement of performance in python-nitrate module and implementation of common caching. This required performance test suite, which was created in the beginning.

The main achievements are implementation of Common caching class in this module, but also Persistent cache and MultiCall feature, which improved performance of updating part of module. Tag class is also introduced and draft of container caching/initialization.

The important results are those created in the testing part that these feature enhancement really improved performance in python-nitrate module.

Bibliography

[1] MOLYNEAUX, I. *The Art of Application Performance Testing*. 1. vyd. Cambridge: O'Reilly Media, 2009. 176 s. ISBN 978-0-596-52066-3.